Exercises of Linux commands

Joseba Hernández Bravo and Jorge Vicente Puig

February 7, 2022

First of all, we have to make some changes to "articles.csv". Fields (columns) in our file are separated by "," -s. However, in some cases (in particular with the titles), we have some "," -s that don't work as a separator. Thus, we have removed (and substituted by "_") and we have created a new file called "articles corrected.csv" with all the changes.

```
sed 's/',[[:space:]]'/_/g' articles.csv | sed 's/'_Cast'/',Cast'/' > articles_corrected.csv
```

1 Question

Take a look at the last 10 lines of the articles file. Which command are you going to use? Modify the command to show just the last line of the file.

We will use the command *tail*.

```
tail -10 articles_corrected.csv
```

```
04 Jan 2017,Article,Seth Kenlon,By Jove! It's a lightweight alternative to Vim,14,/article/17/1/jove-lightweight-alternative-vim,Text editors,1346
04 Jan 2017,Article,Preston Ward,DronePan: An app that captures panorama views with your aircraft,2,/article/17/1/dronepan,Hardware,533
03 Jan 2017,Article,Shawn Powers,4 hot skills for Linux pros in 2017,10,/article/17/1/yearbook-4-hot-skills-linux-pros-2017,"2016 Open Source Yearbook_Yearbook_Careers",708
03 Jan 2017,Article,Anna Ossowski,What does cross stitch have to do with programming? More than you think,5,/article/17/11/traditional-arts-crafts-code-programming,"JavaScript_Programming",1997
03 Jan 2017,Article,Mihai Raulea,Tapitoo OpenCart: An open source e-commerce mobile app,1,/article/17/1/tapitoo-opencart,Mobile,1127
03 Jan 2017,Article,Sam Knuth,Avoid echo chambers and make open decisions,1,/open-organization/17/1/avoid-echo-chambers-make-open-decisions,The Open Organization,723
02 Jan 2017,Article,Richard Fontana,7 notable legal developments in open source in 2016,3,/article/17/1/yearbook-7-notable-legal-developments-2016,"2016 Open Source Yearbook_Yearbook_Licensing",2359
02 Jan 2017,Article,Scott Nesbitt,3 tips for effectively using wikis for documentation,1,/article/17/1/tips-using-wiki-documentation,"Documentation_Wiki",710
02 Jan 2017,Article,Jen Wike Huger,The Opensource.com preview for January,0,/article/17/1/editorial-preview-january,,358
02 Jan 2017,Poll,Jason Baker,What is your open source New Year's resolution?,1,/poll/17/1/what-your-open-source-new-years-resolution,,186
```

```
tail -1 articles_corrected.csv
```

```
02 Jan 2017,Poll,Jason Baker,What is your open source New Year's resolution?,1,/poll/17/1/what-your-open-source-new-years-resolution,,186
```

2 Question

Extract all lines that belong to January 6th from the articles.csv file and store them in a new file named "reyes.csv". Check that the first line of the new file has expected values.

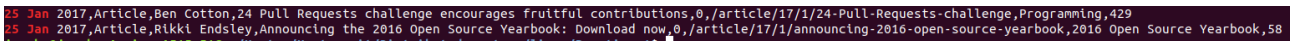
```
grep -i "06 Jan" articles_corrected.cvs > reyes.csv
head -1 reyes.csv
```

```
06 Jan 2017,Article,Jen Wike Huger,"Top 5: Hot programming trends_How Linux got to be Linux_and more",0,/article/17/1/top-5-january-6,Top 5,241
```

3 Question

Use the original csv to find which entries have 0 at the comment count only for those entries from january 25th.

```
awk -F ',' '{if($5 == 0)print $0}' articles.csv | grep -i "25 Jan"
```



4 Question

Now count the number of entries of Q3 and compare it with the total number of entries.

```
grep -i "25 jan" articles.csv | cut -d "," -f5 | grep "0" | wc -l
```

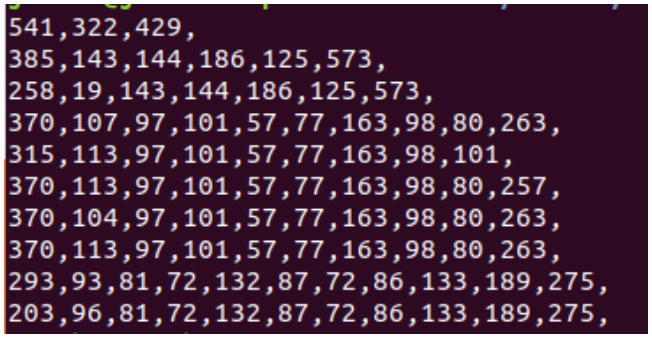
```
wc -l articles.csv
```

In the first case we obtain only 2 entries while we obtain 93 in the total number of entries file.

5 Question

Now use *mRNA*.bed file. In this file, we are interested in the exon sizes of each entry. They are located in field number 11. Now you have to get the exon sizes of the first 10 entries of the file.

```
cut -d '    ' -f11 mRNA.bed | head -10
```

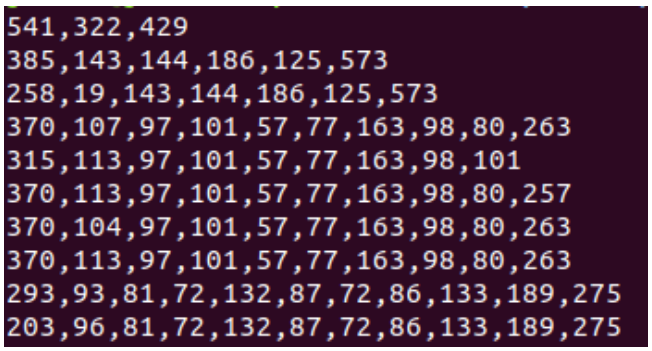


6 Question

How would you remove the last comma?

```
cut -d '    ' -f11 mRNA.bed | head -10 | sed 's/.$//'
```

The dot indicates any character in sed and the $ indicates the end of the line. In other words "$." means, delete the last character only. Result:



7 Question

How would get the smallest size from each of the records? The result should provide a number for each line of the input.

We will use the next reasoning:

- First we extract the 11th column and print 10 first values, remove the last ”,” and save it into ”first_exon_sizes.txt” using the command:

```
cut -d '    ' -f11 mRNA.bed | head -10 | sed 's/.$//' > first_exon_sizes.txt
```

- Then, we replace the delimiter ”,” by ” ” and print the minimum value of each line using the command:

```
sed 's/,/ /g' first_exon_sizes.txt |
awk '{m=$1;for(i=1;i<=NF;i++)if($i<m)m=$i;print "min of line",NR": ",m$}'
```

However we can do everything in the same line by using the command:

```
cut -d '    ' -f11 mRNA.bed | head -10 | sed 's/,/ /g' |
awk '{m=$1;for(i=1;i<=NF;i++)if($i<m)m=$i;print "min of line",NR": ",m$}'
```

With the result:

```
min of line 1:  322
min of line 2:  125
min of line 3:  19
min of line 4:  57
min of line 5:  57
min of line 6:  57
min of line 7:  57
min of line 8:  57
min of line 9:  72
min of line 10:  72
```

8 Question

How would you now sort the records so that the first number shown is the smallest exon size? Again, the answer must provide a sorted list of numbers for each line of the input.

```
cut -d '    ' -f11 mRNA.bed | head -10 | sed 's/,/ /g' |
awk '{m=$1;for(i=1;i<=NF;i++)if($i<m)m=$i;print "min of line",NR": ",m$}' | sort -k5n
```

With the resuts:

```
min of line 3:  19
min of line 4:  57
min of line 5:  57
min of line 6:  57
min of line 7:  57
min of line 8:  57
min of line 10:  72
min of line 9:  72
min of line 2:  125
min of line 1:  322
```

9 Question

Now get the 10 largest exons of chr1 stored in *mRNA.**bed***.

```
grep -i "chr1" mRNA.bed | cut -d '  ' -f11 | sed 's/,/ /g' |
awk '{m=$1;for(i=1;i<=NF;i++)if($i>m)m=$i;print "max of line",NR": ",m}' | sort -k5nr | head -10
```

With the results:

```
max of line 323:  7713
max of line 5224:  5616
max of line 8124:  5239
max of line 2039:  4755
max of line 3559:  4154
max of line 998:  4075
max of line 8445:  3897
max of line 613:  3882
max of line 5439:  3875
max of line 3341:  3757
```

10 Question

Now modify Q9 script to recieve as a parameter the number of exons to search for.

In this exercise we have the program on the script *get_maximum.sh*:

```
#!/bin/bash
Axon_number=$1
echo \We are going to calculate the largest \"$Axon_number\" axons in chr1.
grep -i "chr1" mRNA.bed | cut -d '' -f11 | sed 's/,/ /g' > largest_axons.txt
awk '{m=$1;for(i=1;i<=NF;i++)if($i>m)m=$i;print "max of line",NR": ",m}' largest_axons.txt |
                                                        sort -k5nr | head -$Axon_number
```

For instance we can calculate the 7 largest exons:

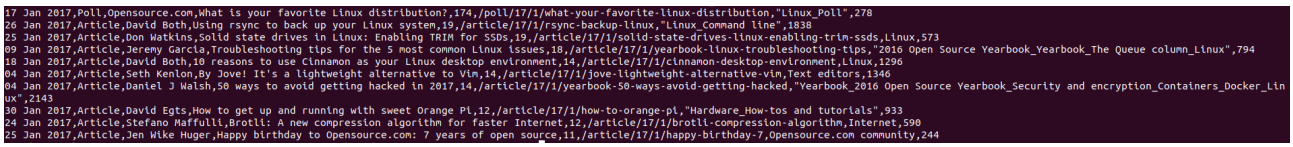
```
./get_maximum.sh 7
```

```
max of line 323:  7713
max of line 5224:  5616
max of line 8124:  5239
max of line 2039:  4755
max of line 3559:  4154
max of line 998:  4075
max of line 8445:  3897
```

11 Question

Get the first 10 records of ***articles.csv*** with largest number of comments from the original csv file.

```
sort -nr -t ',' -k5 articles.csv | head -10
```



12 Question

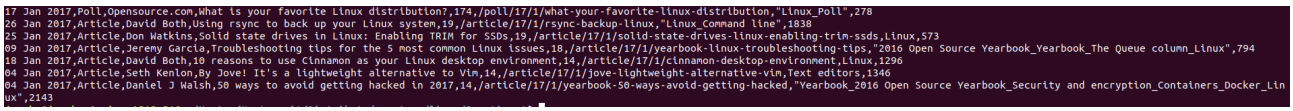
Modify your previous script to recieve a number as a parameter ***N*** and then show the top ***N*** entries with more comments.

We will use the program on the script *topNentries.sh*:

```
#!/bin/bash
N=$1
sort -nr -t ',' -k5 articles_corrected.csv | head -$N
```

And for instance, we can show the 7th top entries with more comments:

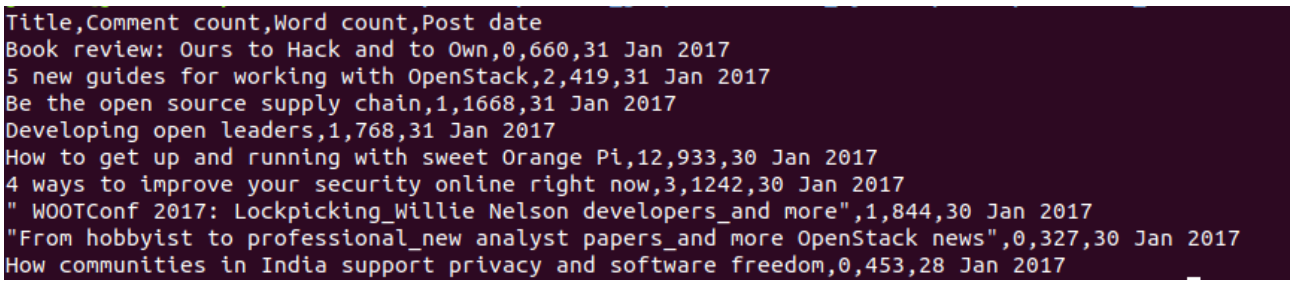
```
./topNentries.sh 7
```



13 Question

Mow we are going to create a new ***articles.csv*** where we get a different output data layout using awk tool. INPUT: Post date, Content type, Author, Title, Comm count, Path, Tags, Word count OUTPUT: Title; Comment count; Word count; Post date

```
awk -F ',' '{print $4","$5","$8","$1}' articles_corrected.csv > articles_Q13.txt
head -10 articles_Q13.txt
```



14 Question

Now create a new article2.csv format where we cut the Title text to 10 characters and we get only the last level of the Path.

```
#!/bin/bash
cut -d ',' -f4 articles_corrected.csv | cut -c1-10 > cut_titles.csv
cut -d ',' -f6 articles_corrected.csv | cut -f5 -d'/' > cut_path.csv
paste cut_titles cut_path > pasted_title_path.csv
head -10 pasted_title_path.csv
```

```
Title   Path
Book revie      review-book-ours-to-hack-and-own
5 new guid      openstack-tutorials
Be the ope      be-open-source-supply-chain
Developing      developing-open-leaders
How to get      how-to-orange-pi
4 ways to       4-ways-improve-your-online-security
" WOOTConf      women-open-technology-miniconf
"From hobb      openstack-news-january-30
How commun      how-communities-india-support-privacy-software-freedom
```